

Deep Learning Aided Visual Localisation in Urban Pedestrian Environments

by Maleen Jayasuriya

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of
Prof. Gamini Dissanayake and Dr. Ravindra Ranasinghe

University of Technology Sydney
Faculty of Engineering and Information Technology

July 2021

Certificate of Original Authorship

I, Maleen Jayasuriya declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program

Signed: Production Note:
Signature removed prior to publication.

Date: 26/07/2021

Deep Learning Aided Visual Localisation in Urban Pedestrian Environments

by

Maleen Jayasuriya

A thesis submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy

Abstract

Localisation of a mobile robot is a fundamental problem in robotics research. In a known environment, localisation can be performed using a prebuilt map, whereas much more complex simultaneous localisation and mapping (SLAM), which estimates both the robot location and the map, is required when operating in an unknown environment. This thesis focuses on the localisation of low-speed vehicles ranging from personal mobility devices to delivery robots, operating in a known outdoor urban environment using low-cost cameras with the objective of improving their functionality and safety. Existing techniques for vision only localisation, even while operating in known environments, requires SLAM due to the difficulty in building reliable maps that are persistent across long time frames.

This thesis proposes an approach that circumvents this problem by utilising convolutional neural network (CNN) based perception of (a) persistent pole-like landmarks such as lamp posts, trees, street signs and parking meters, and (b) important ground surface boundaries related to persistent infrastructure such as curbs, pavement edges and manhole covers, found in urban environments. Localisation is carried out on a prebuilt map consisting of the 2D locations of these landmarks and a vector distance transform (VDT) representation of the ground surface boundaries. An extended Kalman filter (EKF) fuses these observations to carry out pose estimation while robustly dealing with missed detections and wrong classifications.

This approach is further extended by utilising an omnidirectional camera to improve the effective field of view (FoV) of the landmark detection system. The framework utilises an information theoretic strategy to decide the best viewpoint to serve as an input to the CNN in a given iteration, instead of the full 360° coverage offered by an omnidirectional camera, in order to leverage the advantage of having a higher field of view without compromising on performance.

The final contribution of this thesis is a strategy to incorporate the knowledge of traversable paths in the environment into the overall localisation framework, as the target applications predominantly travel on pavements or footpaths. It is demonstrated that enforcing the constraint that the vehicle can only traverse on specific regions in a given map significantly enhances the quality of the location estimate. A decision making framework to ascertain whether the traversability constraint should be enforced or whether there has been a deliberate action to move the vehicle out of the predefined path is also presented.

Real-world experiments carried out in dynamic urban environments across large time gaps in the year and at different distance scales, utilising an instrumented mobility scooter, are presented to highlight the effectiveness of the contributions of this thesis in contrast to state of the art visual SLAM based approaches.

Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Gamini Dissanayake and Dr. Ravindra Ranasinghe for taking me on as their student. Their advice, support and guidance has not only been instrumental in this work, but the example they set with their character, humility and work ethic will remain a constant inspiration.

I would like to express my gratitude to Dr. Janindu Arukgoda who was a constant source of advice and support both in terms of the theoretical as well as implementation aspects of this thesis. I would also like to thank Mr. Josh Olsen and Mr. Nathanael Gandhi whose support was instrumental in the development of the mobility scooter hardware platform. Likewise, the immense help received from Mr. Julien Collart, Mr. Peter Morris, Mr. David Murphy, and Ms. Kavindie Katuwandeniya in conducting experiments is greatly appreciated.

I would like to acknowledge the warmth and friendship of the staff and students at the University of Technology (UTS) Robotics Institute (RI) which made studying and working an absolute joy and privilege. The knowledge sharing sessions, reading groups, talks, and seminars in particular have been a great source of knowledge and inspiration. I am deeply appreciative of the many friendships that made the less glamorous aspects of my research, tolerable. This includes those who have been already mentioned, as well as Dr. Masha Popović, Mr. James Unicomb, Dr. Richardo Khonasty, Ms. Sheila Sutjipto, Mr. Yujun Lai, Mr. Stefano Aldini, Mr. Stefan Kiss, Mr Hongkyoon Byun, Ms. Mitchell Usayiwevu, Mr. Jesse Mehami, Ms. Lan Wu, Ms. Anna Lidfors Lindqvist, and many more who made my time at UTS RI a positive one.

I am also grateful to Dr. Godaliyadda and Dr. Parakarama, my supervisors at the University of Peradeniya for their guidance, support and particularly the advice to follow a career in academia.

Finally I would like to acknowledge the incredible support system made up of my family and friends in Sri Lanka. My brother and sister-in-law for taking care of me in Sydney (including the home cooked Sri Lankan meals). My sister for her support. My fiancée for being a constant rock, creative partner, and friend through the good times and bad. Last but not least to my parents for their immeasurable hard work, love, guidance and support. This thesis is dedicated to them.

Contents

Declaration of Authorship	iii
Abstract	v
Acknowledgements	vii
List of Figures	xiii
List of Tables	xv
Nomenclature	xix
1 Introduction	1
1.1 Motivation and Scope	1
1.2 Thesis	3
1.3 Contributions	5
1.4 Publications	6
1.5 Competitions and Awards	7
1.6 Thesis Outline	7
2 Background, Context and Related Work	9
2.1 Introduction	9
2.2 The Pose Estimation Problem	10
2.3 Sensor Modalities for Outdoor Urban Navigation	14
2.4 Vision-based Pose Estimation	16
3 A CNN based Visual Localisation Strategy for Urban Environments	21
3.1 Introduction	21
3.2 Background	22
3.2.1 Convolutional Neural Networks and Pose Estimation	22
3.2.2 Distance Transforms and Pose Estimation	24
3.3 The UV-Loc Framework	27
3.3.1 Front-End	27
3.3.1.1 Odometry	28

3.3.1.2	Landmark Observations	28
3.3.1.3	Ground Surface Observations	30
3.3.2	Back-End	31
3.3.2.1	Prediction Step	32
3.3.2.2	Update Using Landmark Observations	33
3.3.2.3	Update Using Ground Surface Observations	35
3.4	Hardware Overview	38
3.4.1	Mobility Scooter	38
3.4.2	Vision Sensors	39
3.4.3	Odometry Sensors	40
3.4.4	Computing	40
3.4.5	Sensors for Obtaining the Ground Truth	41
3.5	Map Construction and Determining Sensor Noise	42
3.5.1	Obtaining Known Vehicle Poses	43
3.5.2	Sensor Noise	44
3.6	Experimental Results and Discussion	44
3.6.1	Experiment 01: Validation of UV-Loc at Wentworth Park, Sydney	45
3.6.1.1	Experiment 01: Discussion	47
3.6.2	Experiment 02: Evaluation of UV-Loc at Glebe, Sydney Across Different Time Frames	48
3.6.2.1	Experiment 02: Discussion	50
3.6.3	Experiment 03: Evaluation of UV-Loc on a Large-scale Dataset at Ultimo, Sydney	52
3.6.3.1	Experiment 03: Discussion	54
3.7	Summary	55
4	Active Omnidirectional Perception for Urban Visual Localisation	57
4.1	Introduction	57
4.2	Background	58
4.2.1	Omnidirectional Vision	58
4.2.2	Active Perception	59
4.3	The UV-Loc-360 Framework	60
4.3.1	Overview	60
4.3.2	Omnidirectional Perception	61
4.3.3	CNN Based Feature Extraction for Omnidirectional Images	64
4.3.4	Active Vision	65
4.4	Hardware Modifications	67
4.4.1	Calibration	68
4.5	Experimental Results and Discussion	69
4.5.1	Experiment 01: Validation of UV-Loc-360 at Wentworth Park, Sydney	70
4.5.2	Experiment 02: Evaluation of UV-Loc-360 at Glebe, Sydney Across Different Time Frames	71
4.5.3	Experiment 03: Evaluation of UV-Loc-360 on a Large-scale Dataset at Ultimo, Sydney	73

4.5.4	Discussion	75
4.6	Conclusion	76
5	Enforcing Path Traversability Constraints to Aid Visual Urban Localisation	79
5.1	Introduction	79
5.2	Background	80
5.3	The UV-Loc-constrained Framework	82
5.3.1	Formulation of Road Boundary Constraints	83
5.3.2	Calculating the Constrained Estimate	86
5.3.3	Evaluating the Applicability of Constraints	86
5.4	Experimental Results and Discussion	87
5.4.1	Experiment 01: Validation of UV-Loc-constrained at Wentworth Park, Sydney	88
5.4.1.1	Experiment 01: Discussion	90
5.4.2	Experiment 02: Evaluation of UV-Loc-constrained at Glebe, Sydney Across Different Time Frames	91
5.4.2.1	Experiment 02: Discussion	93
5.4.3	Experiment 03: Evaluation of UV-Loc-constrained on a Large-scale Dataset at Ultimo, Sydney	94
5.4.3.1	Experiment 03: Discussion	95
5.5	Conclusion	96
6	Conclusion	99
6.1	Summary of Datasets and Results	100
6.2	Summary of Contributions and Limitations	103
6.2.1	UV-Loc: A CNN based Visual Localisation Strategy for Urban Environments	103
6.2.2	UV-Loc-360: Active Omnidirectional Perception for Urban Visual Localisation	104
6.2.3	UV-Loc-constrained: Enforcing Path Traversability Constraints to Aid Visual Urban Localisation	104
6.3	Discussion on Future Work	105

List of Figures

3.1	Example closed shaped environment	26
3.2	Variation of SDT, UDT and VDT components along the dotted line in Figure 3.1	26
3.3	The architecture of the UV-Loc framework	27
3.4	YOLO based landmark detection	29
3.5	HED CNN based edge detection	30
3.6	Example landmark map (Glebe, Sydney)	33
3.7	Example VDT based ground surface map (Glebe, Sydney)	36
3.8	Hardware overview of the retrofitted mobility scooter	39
3.9	Trajectory of UV-Loc and state of the art systems at Wentworth Park, Sydney relative to RTK-GNSS ground truth for validation and tuning . . .	46
3.10	UV-Loc estimation error at Wentworth Park relative to RTK-GNSS ground truth (red), with 2σ covariance bounds (blue)	46
3.11	Trajectory of UV-Loc and state of the art systems at Glebe, Sydney relative to sporadic RTK-GNSS ground truth, to evaluate performance across different time frames	49
3.12	UV-Loc estimation error at Glebe, Sydney relative to sporadic RTK-GNSS ground truth (red), with 2σ covariance bounds (blue)	49
3.13	Locations in Glebe six months apart	51
3.14	Trajectory of UV-Loc and state of the art systems at Ultimo, Sydney relative to RTAB-MAP laser based ground truth, to evaluate performance at large scales	52
3.15	UV-Loc estimation error at Ultimo, Sydney relative to RTAB-MAP laser based ground truth (red), with 2σ covariance bounds (blue)	53
3.16	Examples of litter and debris on ground surface	55
4.1	UV-Loc-360 Architecture	61
4.2	Spherical projection model	62
4.3	Projection Pipeline	63
4.4	Virtual perspective camera viewpoints	66
4.5	Hardware overview of retrofitted mobility scooter	67
4.6	Extrinsic calibration of a single viewpoint	69
4.7	Trajectory of UV-Loc-360 in comparison to UV-Loc, at Wentworth Park, Sydney relative to RTK-GNSS ground truth for validation and tuning . . .	70
4.8	UV-Loc-360 estimation error at Wentworth Park relative to RTK-GNSS ground truth (red), with 2σ covariance bounds	71

4.9	Trajectory of UV-Loc-360 in comparison to UV-Loc at Glebe, Sydney relative to sporadic RTK-GNSS ground truth, to evaluate performance across different time frames	72
4.10	UV-Loc-360 estimation error at Glebe, Sydney relative to sporadic RTK-GNSS ground truth (red), with 2σ covariance bounds (blue)	72
4.11	Trajectory of UV-Loc-360 in comparison to UV-Loc at Ultimo, Sydney relative to RTAB-MAP laser based ground truth, to evaluate performance at large scales	74
4.12	UV-Loc-360 estimation error at Ultimo, Sydney relative to RTAB-MAP laser based ground truth (red), with 2σ covariance bounds (blue)	74
5.1	The UV-Loc-constrained architecture	83
5.2	Example Polyline map for Glebe, Sydney	84
5.3	Constraints formulation	85
5.4	Localisation results at Wentworth Park, Sydney	88
5.5	Estimation error for trajectory 01 (red), with 2σ covariance bounds (blue)	89
5.6	Trajectory of UV-Loc-constrained in comparison to UV-Loc at Glebe, Sydney relative to sporadic RTK-GNSS ground truth, to evaluate performance across different time frames	91
5.7	UV-Loc-constrained estimation error at Glebe, Sydney relative to sporadic RTK-GNSS ground truth (red), with 2σ covariance bounds (blue)	92
5.8	Example zoomed in sections of trajectories in Glebe, Sydney before and after applying constraints	93
5.9	Trajectory of UV-Loc-constrained in comparison to UV-Loc at Ultimo, Sydney relative to RTAB-MAP laser based ground truth, to evaluate performance at large scales	94
5.10	UV-Loc-constrained estimation error at Ultimo, Sydney relative to RTAB-MAP laser based ground truth (red), with 2σ covariance bounds (blue)	95
5.11	Example zoomed in sections of trajectories in Ultimo, Sydney before and after applying constraints	96
6.1	Summary of Localisation Trajectories of Visual Odometry, UV-Loc, UV-Loc-360, and UV-Loc-constrained	102

List of Tables

3.1	Pathrider 10 Physical Specifications	39
3.2	Computational performance on the hardware platform	41
3.3	Experiment 01: Root Mean Square Errors for UV-Loc and state of the art systems at Wentworth Park, Sydney relative to RTK-GNSS ground truth . .	47
3.4	Experiment 02: Root Mean Square Errors for UV-Loc and state of the art systems at Glebe, Sydney relative to sporadic RTK-GNSS ground truth . .	50
3.5	Experiment 03: Root Mean Square Errors for UV-Loc and state of the art systems at Ultimo, Sydney relative to RTAB-MAP laser based ground truth	53
4.1	Experiment 01: Root Mean Square Errors for UV-Loc-360, UV-Loc and state of the art systems at Wentworth Park, Sydney relative to RTK-GNSS ground truth	71
4.2	Experiment 02: Root Mean Square Errors for UV-Loc-360, UV-Loc and state of the art systems at Glebe, Sydney relative to sporadic RTK-GNSS ground truth	73
4.3	Experiment 03: Root Mean Square Errors for UV-Loc-360, UV-Loc and state of the art systems at Ultimo, Sydney relative to RTAB-MAP laser based ground truth	75
5.1	Experiment 01: Root Mean Square Errors for trajectory 01	89
5.2	Experiment 01: Root Mean Square Errors for trajectory 2 to 4 [m]	90
5.3	Experiment 02: Root Mean Square Errors for UV-Loc-constrained and state of the art systems at Glebe, Sydney relative to sporadic RTK-GNSS ground truth	92
5.4	Experiment 03: Root Mean Square Errors for UV-Loc-constrained, and state of the art systems at Ultimo, Sydney relative to RTAB-MAP laser based ground truth	95
6.1	Summary of Datasets and Results	101

Acronyms & Abbreviations

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
UTS	University of Technology Sydney
RI	Robotics Institute
PMD	Personal Mobility Device
HMI	Human-Machine Interface
GNSS	Global Navigation Satellite Systems
LIDAR	Light Detection and Ranging
RADAR	Radio Detection and Ranging
SONAR	Sound Navigation and Ranging
SLAM	Simultaneous Localisation and Mapping
CNN	Convolutional Neural Network
EKF	Extended Kalman Filter
DOF	Degrees of Freedom
GPS	Global Positioning System
IMU	Inertial Measurement Unit

RTK-GNSS	Real-time Kinematics GNSS
RTK-GNSS	Differential GNSS
RTK-GNSS	Assisted GNSS
RGB-D	Red, Green, Blue and Depth
BA	Bundle Adjustment
RTAB	Real-Time Appearance-Based Mapping
HED	Holistically Nested Edge Detection
YOLO	You Only Look Once
SDT	Signed Distance Transform
UDT	Unsigned Distance Transform
VDT	Vector Distance Transform
VO	Visual Odometry
OGM	Occupancy Grid Map
FoV	Field of View
UV-Loc	Urban Visual Localiser
RMSE	Root Mean Square Error

Nomenclature

General Notations

t	Time (continuous)
k	Time (discrete step)
v	Linear velocity
ω	Angular velocity
U_t	A vector containing the linear and angular velocity at time t
θ_t^i	i^{th} bearing measurement to landmark observation at time t
Θ_t	A vector containing all bearing measurement to landmark observation at time t
l_t^i	i^{th} semantic label to landmark observation at time t
L_t	A vector containing all semantic labels of landmark observations at time t
M_L	Environmental landmark map
$(x_{z,t}^i, y_{z,t}^i)$	i^{th} ground surface boundary point at time t
Z_t	A vector ground surface boundary points at time t
DT_v	Vector Distance Transform
DT_x	x component of the Vector Distance Transform
DT_y	y component of the Vector Distance Transform
M_G	Ground surface boundary map

$\hat{\mathbf{X}}_t$	Pose estimate of robot in 2D space at time t . Consists of position components \hat{x}_t , \hat{y}_t and the orientation component $\hat{\phi}_t$.
P_t	Robot pose covariance matrix at time t
\square_{t-1}	Previous state
$\square_{t t-1}$	Predicted current state
\square_t	Updated current state
$g(\cdot, \cdot)$	Motion model
$h(\cdot, \cdot)$	Observation model
∇G_{\square}	Jacobian of the motion model with respect to \square
∇H_{\square}	Jacobian of the observation model with respect to \square
ν	Innovation vector
S	Innovation covariance
K	Kalman gain
\square_l	EKF Update parameters related to landmark observations
\square_g	EKF Update parameters related to ground surface observations
vp	A set of perspective camera viewpoints obtained from an omnidirectional camera
ΔP^{vp}	A derived information metric that quantifies the impact on the overall pose uncertainty from a camera viewpoint
${}^A\mathbf{T}_B$	The transform matrix of frame B relative to frame A
\square_u	Refers to unconstrained states, covariances and related parameters
\square_c	Refers to constrained states, covariances and related parameters
${}^{nt}\square$	Pose components in local tangent frame
M_P	Polyline boundary map
ub_t	Upper path boundary at time t
lb_t	Lower path boundary at time t
d^2	Mahalanobis distance

$cov(\cdot)$	Covariance matrix
$diag(\cdot)$	Diagonal matrix
$trace(\cdot)$	Trace of a matrix
$\Sigma(\cdot)$	Summation
\int	Integral operator
∂	Partial differentiation operator

